



## wwPDB EM Map/Model Validation Report ⓘ

Apr 10, 2016 – 01:52 PM BST

PDB ID : 3J9B  
Title : Electron cryo-microscopy of an RNA polymerase  
Authors : Chang, S.H.; Sun, D.P.; Liang, H.H.; Wang, J.; Li, J.; Guo, L.; Wang, X.L.; Guan, C.C.; Boruah, B.M.; Yuan, L.M.; Feng, F.; Yang, M.R.; Wojdyla, J.; Wang, J.W.; Wang, M.T.; Wang, H.W.; Liu, Y.F.  
Deposited on : 2014-12-16  
Resolution : 4.30 Å(reported)  
Based on PDB ID : 3CM8

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.  
For rigid body fitted models, validation errors reported here could stem from errors in the original structure(s) used in the fitting.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/EMValidationReportHelp>

---

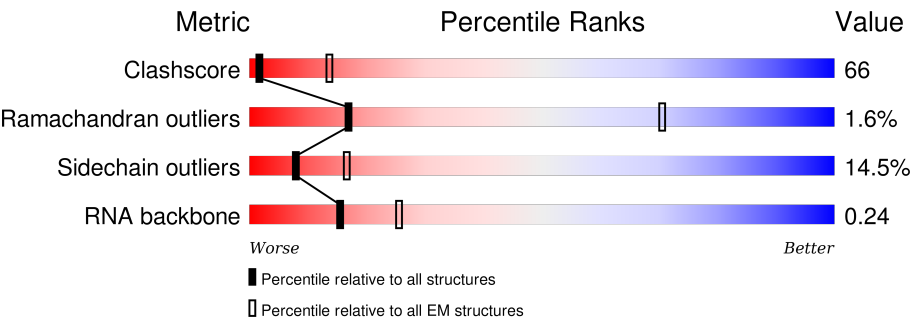
MolProbity : 4.02b-467  
Mogul : unknown  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk27241

# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:  
*ELECTRON MICROSCOPY*

The reported resolution of this entry is 4.30 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.







Metric	Whole archive (#Entries)	EM structures (#Entries)
Clashscore	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686
RNA backbone	3027	244

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	509	<div><div>39%</div><div>47%</div><div>6%</div><div>7%</div></div>
1	H	509	<div><div>39%</div><div>47%</div><div>6%</div><div>7%</div></div>
2	B	440	<div><div>75%</div><div>24%</div><div>.</div></div>
2	I	440	<div><div>75%</div><div>24%</div><div>.</div></div>
3	C	66	<div><div>86%</div><div>14%</div></div>
3	J	66	<div><div>86%</div><div>14%</div></div>
4	D	6	<div><div>83%</div><div>17%</div></div>
4	K	6	<div><div>67%</div><div>17%</div><div>17%</div></div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain	
5	E	6	 83%	 17%
6	L	6	 83%	 17%

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 12854 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Polymerase.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	471	Total	C	N	O	S	0	0
			3618	2301	622	669	26		
1	H	471	Total	C	N	O	S	0	0
			3618	2301	622	669	26		

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	553	GLU	ALA	CONFLICT	UNP Q91R78
A	567	THR	ALA	CONFLICT	UNP Q91R78
H	553	GLU	ALA	CONFLICT	UNP Q91R78
H	567	THR	ALA	CONFLICT	UNP Q91R78

- Molecule 2 is a protein called RNA-directed RNA polymerase catalytic subunit.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	440	Total	C	N	O	S	0	0
			2268	1373	444	450	1		
2	I	440	Total	C	N	O	S	0	0
			2268	1373	444	450	1		

- Molecule 3 is a protein called Polymerase basic protein 2.

Mol	Chain	Residues	Atoms				AltConf	Trace
3	C	66	Total	C	N	O	0	0
			330	198	66	66		
3	J	66	Total	C	N	O	0	0
			330	198	66	66		

- Molecule 4 is a RNA chain called RNA (5'-R(P\*AP\*AP\*AP\*AP\*AP\*A)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	5	Total	C	N	O	P	0	0
			110	50	25	30	5		
4	K	5	Total	C	N	O	P	0	0
			110	50	25	30	5		

- Molecule 5 is a RNA chain called RNA (5'-R(\*UP\*UP\*UP\*UP\*UP\*A)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
5	E	5	Total	C	N	O	P	0	0
			102	46	13	38	5		

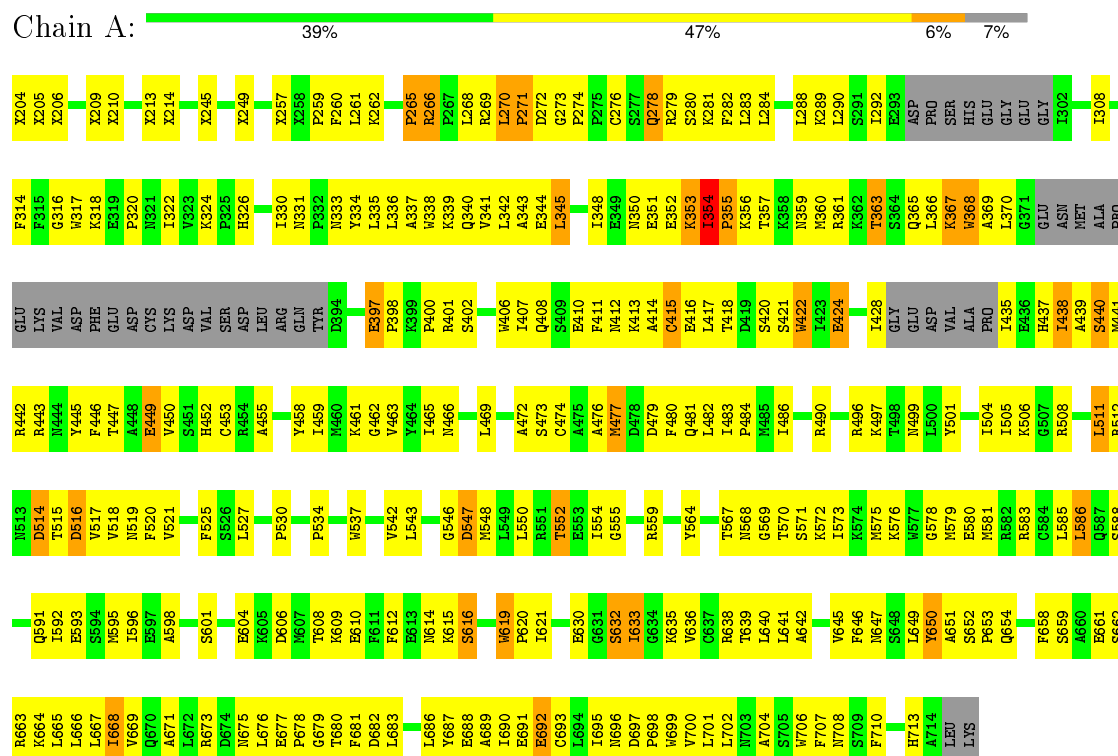
- Molecule 6 is a RNA chain called RNA (5'-R(P\*UP\*UP\*UP\*UP\*UP\*U)-3').

Mol	Chain	Residues	Atoms					AltConf	Trace
6	L	5	Total	C	N	O	P	0	0
			100	45	10	40	5		

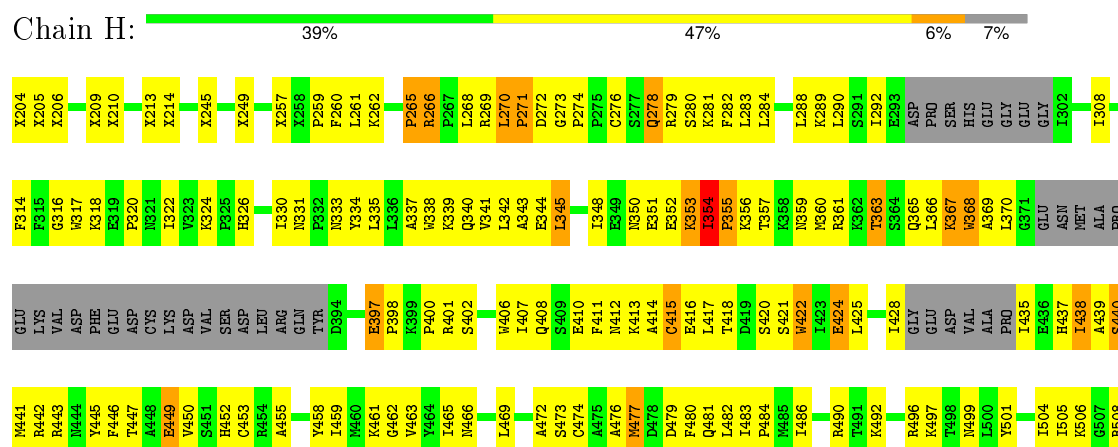
### 3 Residue-property plots

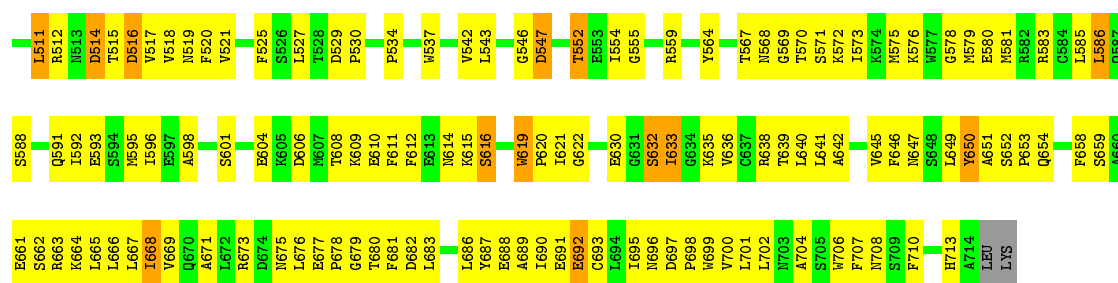
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

#### • Molecule 1: Polymerase



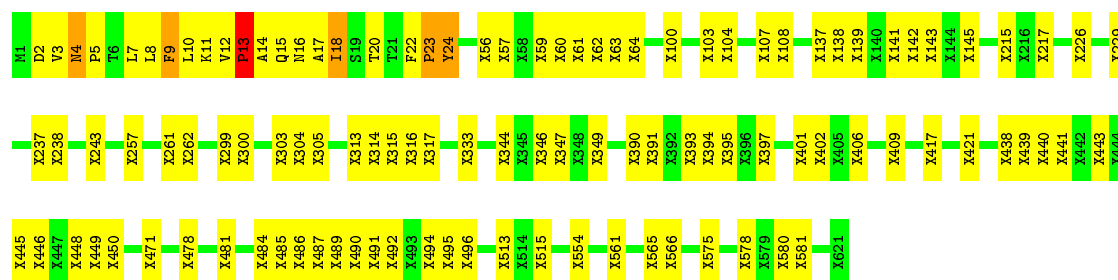
#### • Molecule 1: Polymerase





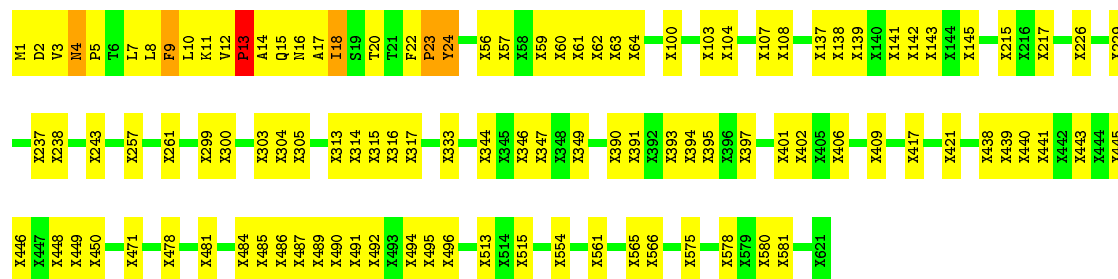
• Molecule 2: RNA-directed RNA polymerase catalytic subunit

Chain B:  75% 24%




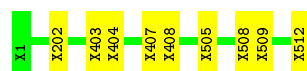
• Molecule 2: RNA-directed RNA polymerase catalytic subunit

Chain I:  75% 24%




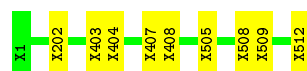
• Molecule 3: Polymerase basic protein 2

Chain C:  86% 14%




• Molecule 3: Polymerase basic protein 2

Chain J:  86% 14%



• Molecule 4: RNA (5'-R(P\*AP\*AP\*AP\*AP\*A)-3')

Chain D:  83% 17%

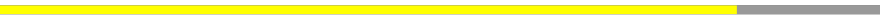
A10  
A11  
A12  
A13  
A14  
A

- Molecule 4: RNA (5'-R(P\*AP\*AP\*AP\*AP\*AP\*A)-3')

Chain K:  67% 17% 17%


A10  
A11  
A12  
A13  
A14  
A

- Molecule 5: RNA (5'-R(\*UP\*UP\*UP\*UP\*UP\*A)-3')

Chain E:  83% 17%

U1  
U2  
U3  
U4  
U  
A6

- Molecule 6: RNA (5'-R(P\*UP\*UP\*UP\*UP\*UP\*U)-3')

Chain L:  83% 17%

U  
U2  
U3  
U4  
U5  
U6



## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of images	67066	Depositor
Resolution determination method	FSC 0.143	Depositor
CTF correction method	Not provided	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	50	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	22500	Depositor
Image detector	GATAN K2 (4k x 4k)	Depositor

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# $ Z  > 2$	RMSZ	# $ Z  > 2$
1	A	0.41	3/3434 (0.1%)	0.59	5/4629 (0.1%)
1	H	0.41	3/3434 (0.1%)	0.59	5/4629 (0.1%)
2	B	0.72	2/193 (1.0%)	0.98	2/265 (0.8%)
2	I	0.72	2/193 (1.0%)	0.98	2/265 (0.8%)
4	D	0.14	0/124	0.58	0/191
4	K	0.14	0/124	0.58	0/191
5	E	0.17	0/112	0.62	0/171
6	L	1.65	5/109 (4.6%)	2.27	8/166 (4.8%)
All	All	0.46	15/7723 (0.2%)	0.68	22/10507 (0.2%)

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	L	6	U	C1'-N1	6.95	1.59	1.48
6	L	5	U	C1'-N1	6.93	1.59	1.48
6	L	2	U	C1'-N1	6.90	1.59	1.48
6	L	3	U	C1'-N1	6.90	1.59	1.48
6	L	4	U	C1'-N1	6.89	1.59	1.48

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	I	13	PRO	CA-N-CD	-8.07	100.21	111.50
2	B	13	PRO	CA-N-CD	-8.06	100.22	111.50
6	L	4	U	OP2-P-O3'	7.21	121.07	105.20
6	L	3	U	OP2-P-O3'	7.21	121.06	105.20
6	L	2	U	OP2-P-O3'	7.20	121.04	105.20

There are no chirality outliers.

There are no planarity outliers.

## 5.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3618	0	3435	572	0
1	H	3618	0	3435	575	0
2	B	2268	0	676	155	0
2	I	2268	0	676	154	0
3	C	330	0	79	8	0
3	J	330	0	79	8	0
4	D	110	0	56	27	0
4	K	110	0	56	25	0
5	E	102	0	52	15	0
6	L	100	0	51	12	0
All	All	12854	0	8595	1418	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 66.

The worst 5 of 1418 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:424:GLU:HB2	1:A:490:ARG:CZ	1.51	1.39
1:H:424:GLU:HB2	1:H:490:ARG:CZ	1.51	1.39
1:A:289:LYS:HG2	1:A:501:TYR:CE1	1.61	1.34
1:H:289:LYS:HG2	1:H:501:TYR:CE1	1.61	1.33
2:I:304:UNK:O	2:I:446:UNK:HA	1.22	1.32

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the backbone conformation was

analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	413/509 (81%)	387 (94%)	22 (5%)	4 (1%)	19	65
1	H	413/509 (81%)	387 (94%)	22 (5%)	4 (1%)	19	65
2	B	22/440 (5%)	19 (86%)	0	3 (14%)	0	6
2	I	22/440 (5%)	19 (86%)	0	3 (14%)	0	6
All	All	870/1898 (46%)	812 (93%)	44 (5%)	14 (2%)	17	57

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	13	PRO
2	I	13	PRO
1	A	265	PRO
1	H	265	PRO
1	A	355	PRO

### 5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all EM entries.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	371/411 (90%)	319 (86%)	52 (14%)	4	28
1	H	371/411 (90%)	319 (86%)	52 (14%)	4	28
2	B	22/22 (100%)	17 (77%)	5 (23%)	1	9
2	I	22/22 (100%)	17 (77%)	5 (23%)	1	9
All	All	786/866 (91%)	672 (86%)	114 (14%)	8	26

5 of 114 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	708	ASN
1	H	326	HIS
1	H	668	ILE
2	B	13	PRO

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	H	260	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	278	GLN
2	B	16	ASN
1	H	278	GLN
1	H	510	HIS
2	I	16	ASN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	D	4/6 (66%)	0	0
4	K	4/6 (66%)	1 (25%)	0
5	E	4/6 (66%)	0	0
6	L	4/6 (66%)	0	0
All	All	16/24 (66%)	1 (6%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
4	K	11	A

There are no RNA pucker outliers to report.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.